WHAT IS CLAIMED IS:

1. A method for analyzing a biological sample comprising converting a first profile of a plurality of measurements of cellular constituents in said biological sample into a projected profile containing a plurality of cellular constituent set values according to a definition of co-varying basis cellular constituent sets, wherein said definition is based upon the co-variation of said cellular constituents under a plurality of different perturbations, and wherein said converting comprises projecting said first profile onto said basis cellular constituent sets.

2. The method of claim 1, wherein the plurality of different perturbations comprises at least five different perturbations.

3. The method of claim 2, wherein the plurality of different perturbations comprises more than ten different perturbations.

4. The method of claim 3, wherein the plurality of different perturbations comprises more than 50 different perturbations.

5. The method of claim 4, wherein the plurality of different perturbations comprises more than 100 different perturbations.

6. The method of claim 1 further comprising the step of indicating the state of said biological sample with said projected profile.

7. The method of claim 1 further comprising the steps of comparing said projected profile with a reference projected profile, and indicating similarity or difference between said projected profile and said reference profile.

8. The method of claim 1, wherein said definition is based upon the co-variation of said cellular constituents under a plurality of different perturbations.

9. The method of claim 8 wherein said definition is defined by a similarity tree derived by a cluster analysis of said cellular constituents under said plurality of perturbations.

10. The method of claim 9 wherein said cellular constituent sets are defined as branches of said similarity tree.

11. The method of claim 10 wherein said branches are selected by applying a cutting level across said tree, wherein said cutting level is determined by expected number of biological pathways represented by said cellular constituents.

12. The method of claim 10 wherein distinction among said branches achieves a statistical significance at 95% confidence level.

13. The method of claim 12 wherein said statistical significance is evaluated with a test using Monte Carlo randomization of an index of said perturbations.

14. The method of claim 13 wherein the test using Monte Carlo randomization comprises:

(a) determining an actual fractional improvement in cluster analysis of said cellular constituents;

(b) generating permuted cellular constituents by means of Monte Carlo randomization of each perturbation for each cellular constituent;

(c) performing cluster analysis on the permuted cellular constituents;

(d) determining the fractional improvements in the cluster analysis of the permuted cellular constituents; and

(e) repeating said steps of generating permuted cellular constituents and performing cluster analysis on the permuted cellular constituents so that a distribution of fractional improvements is obtained,

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

15. The method of claim 12 wherein said statistical significance is evaluated with a test using Monte Carlo randomization of a time index of a biological response to one or more perturbations.

16. The method of claim 10, 11, or 12, wherein said defined cellular constituent sets are refined based upon biological relationships among said cellular constituents.

17. The method of claim 1 wherein said definition is:

$$V = \begin{bmatrix} V_1^{r(1)} & . & V_1^{r(n)} \\ . & . & . \\ V_k^{r(1)} & . & V_k^{r(n)} \end{bmatrix}$$

wherein $V^{(n)}_k$ is the contribution of cellular constituent k to cellular constituent set n.

18. The method of claim 17 wherein said step of converting comprises the execution of the operation:

$$P = [P_1, .. P_i ... P_n] = p \bullet V$$

wherein P_i is cellular constituent set value *i* and vector p is a profile of cellular constituents.

19. The method of claim 1 wherein each of said cellular constituent set values is the average value of the level of said cellular constituents within a corresponding cellular constituent set.

20. The method of claim 1 wherein each of said cellular constituent set value is a weighted average of the level of said cellular constituents within a corresponding cellular constituent set.

21. The method of claim 1 wherein said plurality of measurements is normalized to a unity vector size.

22. The method of claim 1 wherein said measurements of cellular constituents are measurements of responses of said biological sample to a perturbation.

23. A method for analyzing a biological sample comprising:

(a) converting a first profile of a plurality of measurements of cellular constituents in said biological sample into a projected profile containing a plurality of cellular constituent set values according to a definition of co-varying basis cellular constituent sets, wherein said converting comprises projecting said first profile onto said basis cellular constituent sets;

(b) comparing said projected profile with a reference profile; and

(c) indicating similarity or difference between said projected profile and said reference profile.

24. The method of claim 23 wherein said definition is derived from the co-regulation of said cellular constituents.

25. The method of claim 23 wherein said definition is based upon the co-variation of said cellular constituents under a plurality of different perturbations.

26. The method of claim 23 wherein said definition is:

$$V = \begin{bmatrix} V_1^{r(1)} & . & V_1^{r(n)} \\ . & . & . \\ V_k^{r(1)} & . & V_k^{r(n)} \end{bmatrix}$$

wherein $V^{(n)}_k$ is the contribution of cellular constituent k to cellular constituent set n.

27. The method of claim 26 wherein said step of converting comprises the execution of the operation:

$$P = [P_1, .. P_i, .. P_n] = p \bullet V$$

wherein P_i is cellular constituent set value i and vector p is a profile of cellular constituents.

28. The method of claim 23 wherein each of said cellular constituent set values is the average value of the level of said cellular constituents within a corresponding cellular constituent set.

29. The method of claim 23 wherein each of said cellular constituent set value is a weighted average of the level of said cellular constituents within a corresponding cellular constituent set.

30. The method of claim 23 wherein said plurality of measurements is normalized to a unity vector size.

31. The method of claim 23 wherein said measurements of cellular constituents are measurements of responses of said biological sample to a perturbation.

32. A method for analyzing a biological sample comprising converting a first profile of a plurality of measurements of cellular constituents in said biological sample into a projected profile containing a plurality of cellular constituent set values according to a definition of co-varying basis cellular constituent sets,
wherein said definition is provided by the expression

$$V = \begin{bmatrix} V_1^{r(1)} & \cdot & V_1^{r(n)} \\ \cdot & \cdot & \cdot \\ V_k^{r(1)} & \cdot & V_k^{r(n)} \end{bmatrix}$$

in which $V^{(n)}_k$ is the contribution of cellular constituent k to cellular constituent set n, and wherein said converting comprises projecting said first profile onto said basis cellular constituent sets.

33. The method of claim 32 wherein said step of converting comprises the execution of the operation:

$$P = [P_1, .. P_i, .. P_n] = p \bullet V$$

wherein P_i is cellular constituent set value *i* and vector p is a profile of cellular constituents.

34. A method for analyzing a biological sample comprising converting a first profile of a lurality of measurements of cellular constituents in said biological sample into a projected profile containing a plurality of cellular constituent set values according to a definition of co-varying basis cellular constituent sets, each of said cellular constituent set values being a weighted average of the level of said cellular constituent within a corresponding cellular constituent set, wherein said converting comprises projecting said first profile onto said basis cellular constituent sets.

35. A method for analyzing a biological sample comprising converting a first profile of a plurality of measurement of cellular constituents in a biological sample into a projected profile containing a plurality of cellular constituent set values according to a definition of co-varying basis cellular constituent sets, said plurality of measurements being normalized to a unity vector size, wherein said converting comprises projecting said first profile onto said basis cellular constituent sets.

36. A method of grouping biological response profiles according to the similarity of the responses, said method comprising defining similar response profile sets based upon the similarity of a plurality of measured cellular constituents in said response profiles.

37. The method of claim 36, further comprising the step of forming a clustering tree derived by a cluster analysis of similarity of the plurality of measured cellular constituents in said response profiles.

38. The method of claim 37, wherein groups of said biological response profiles are defined as branches of said clustering tree.

39. The method of claim 36, further comprising determining a statistical significance of the groups of biological response profiles.

40. The method of claim 39, wherein the statistical significance of the groups of biological response profiles is determined by means of an objective statistical test.

41. The method of claim 40, wherein the objective statistical test comprises:

(a) determining an actual fractional improvement in cluster analysis of the biological response profiles;

(b) generating permuted response profiles by means of Monte Carlo randomization of each cellular constituent for each response profile;

(c) performing cluster analysis on the permuted response profiles;

(d) determining the fractional improvement in the cluster analysis of the permuted response profiles; and

(e) repeating said steps of generating permuted response profiles and performing cluster analysis on the permuted response profiles so that a distribution of fractional improvements is obtained,

wherein the statistical significance is determined by comparing the actual fractional improvement to the distribution of fractional improvements.

42. A method for analyzing a biological sample comprising:

(a) grouping cellular constituents from the biological sample into sets of cellular constituents that co-vary in biological profiles obtained from the biological sample; and

(b) grouping the biological profiles obtained from the biological sample into sets of biological profiles that effect similar cellular constituents.

43. The method of claim 42, wherein one or more cellular constituents associated with a particular biological effect are identified from said sets of cellular constituents.

44. The method of claim 42, wherein one or more biological profiles associated with a particular biological effect are identified from said sets of biological profiles.

45. The method of claim 43 or 44, wherein the particular biological effect is a biological pathway.

46. The method of claim 43, wherein the cellular constituents from the biological sample comprise a plurality of genes, and one or more genes associated with a particular biological effect are identified.

47. The method of claim 46, wherein the one or more genes identified comprise known genes.

48. The method of claim 46, wherein the one or more genes identified comprise previously unknown genes.

49. The method of claim 42, wherein one or more perturbations associated with a particular biological effect are identified from said sets of biological profiles.

50. The method of claim 49, wherein the one or more perturbations comprise a drug or a drug candidate.

51. The method of claim 50, wherein the one or more perturbations comprise a genetic mutation.

52. The method of claim 50 wherein the drug or drug candidate is a known drug or drug candidate.

53. The method of claim 51, wherein the genetic mutation is a known genetic mutation.

54. The method of claim 50, wherein the drug or drug candidate is a previously unknown drug or drug candidate.

55. The method of claim 51, wherein the genetic mutation is a previously unknown genetic mutation.

56. A method for analyzing an N-dimensional array of data, N being a positive integer, wherein each element of the N-dimensional array of data has N indices, said method comprising grouping each index into sets of data that co-vary within the N-dimensional array of data.

57.. The method of claim 56, wherein each of said sets is defined by a similarity tree derived by a cluster analysis of each of said indices.

58. A method for removing one or more artifacts from a measured biological profile comprising a plurality of measurements of cellular constituents, said method comprising subtracting one or more artifact patterns from the measured biological profile, wherein each of said one or more artifact patterns corresponds to a particular artifact.

59. The method of claim 58, wherein the each of the one or more artifact patterns is provided by knowledge of the genes and relative amplitutdes of responses associated with particular artifact to which each of the one or more artifact patterns corresponds.

60. The method of claim 58, wherein each of the one or more artifact patterns is provided by experiments with perturbations of suspected causative variables of the particular artifact to which each of the one or more artifact patterns corresponds.

61. The method of claim 58, wherein each of the one or more artifact patterns is provided by a cluster analysis of control biological profiles, the control biological profiles comprising a plurality of measurements of cellular constituents in experiments wherein the artifact to which each of the one or more artifact pattern corresponds arises.

62. The method of claim 58, wherein of the one or more artifact patterns are scaled by scaling coefficients, each of the one or more artifact patterns having a particular scaling coefficient.

63. The method of claim 62, wherein the scaling coefficients are determined by a method comprising determining the value of each particular scaling coefficient which minimizes the value of an objective function of the difference between the measured profile and the sum of the one or more scaled artifact patterns.

64. The method of claim 63, wherein the objective function is a least squares minimization.

65. The method of claim 58, wherein each of the one or more artifact patterns is selected from a library of artifact signatures, said artifact signatures corresponding to levels of severity of each the one or more artifacts.

66. The method of claim 65, wherein the artifact signatures are selected by a method comprising determining the artifact signatures which minimize the values of an objective function of the difference between the measured profile and the sum of the one or more artifact signatures.

67. The method of claim 1, wherein the plurality of different perturbations comprises a plurality of graded levels of exposure to a particular perturbation.

68. The method of claim 67, wherein the particular perturbation is a drug or drug candidate.

69. The method of claim 1, wherein said definition is based upon the co-variation of the cellular constituents over a period of time.

add a7